VEGETATION CLASSIFICATION AND SURVEY

International Association for Vegetation Science (IAVS)

Vegetation Classification and Survey 5: 169–177
doi: 10.3897/VCS.126956

FORUM PAPER

Determinants of citation impact

Jürgen Dengler[1,2]

1 *Vegetation Ecology Research Group, Institute of Natural Resource Sciences (IUNR), Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland*
2 *Plant Ecology, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany*

Corresponding author: Jürgen Dengler (juergen.dengler@zhaw.ch)

Academic editor: John Hunter ♦ **Linguistic editor:** Stephen Bell
Received 6 May 2024 ♦ **Accepted** 26 July 2024 ♦ **Published** 23 August 2024

Abstract

This article aims to quantitatively assess how different formal aspects – beyond the relevance and quality of a study – influence how often a scientific paper is cited. As a case study, I retrieved all publications co-authored by myself from the Scopus database, of which 174 could be used for regression modelling. The citation impact was quantified as Field-Weighted Citation Impact (FWCI), which is the citation number normalised by year, subject area and article type. I examined 13 easily accessible numeric and binary predictor variables, including the Source Normalized Impact per Paper (SNIP), open access, special feature, number of authors, length of article and title, as well as formal aspects of the title. In the minimal adequate model, these formal aspects explained 50.2% of the variance in FWCI, with the SNIP alone explaining only 26.8%. Other strong positive predictors were title brevity, article length, special feature and the use of a colon in the title. By contrast, open access and the formulation of titles as factual statements did not have a significant effect. For authors who wish to make their articles more impactful, the main recommendation is to shorten the title and to disregard using factual statements that make the title longer.

Abbreviations: FWCI = Field-weighted Citation Impact; JIF = Journal Impact Factor; OA = open access; SNIP = Source Normalized Impact per Paper: VCS = Vegetation Classification and Survey.

Keywords

article impact, article title, bibliometrics, citation rate, Field-Weighted Citation Impact (FWCI), normalised citation rate, open access, research assessment, Scopus database, special feature, vegetation ecology, Web of Science

Introduction

Authors of scientific papers normally want to achieve impact with their publications, and likewise editors of scientific journals want the published articles to be as impactful as possible. Therefore, the big question is "what makes a paper successful?" Admittedly, the scientific impact of a paper depends on the content, such as the relevance of the topic, state-of-the art techniques in the analyses and well-founded conclusions. Secondly, one would think that the writing style and the appeal of the figures play a role. Both are doubtlessly true, and it is hard to give generic advice on the first point while the second is nicely addressed in various textbooks on scientific writing (Gustavii 2008; Cargill and O'Connor 2009). Moreover, both groups of factors are so diverse that they could hardly be analysed quantitatively.

However, there is a third group of factors that should not be underestimated. These are formal aspects, such as the choice of the journal and of the language, the style of the title and the length of the article. Authors and editors alike invest considerable efforts here. However, there is a lack of empirical studies that test which measures might be effective and to which degree they contribute to the success of a paper. My long-standing impression as co-author and editor is that this field is dominated by either ignorance or strong beliefs, but hardly by empirical facts. To fill this gap, I conducted a quantitative study on how different "formal aspects" influence the citation impact of articles.

PENSOFT

Copyright *Jürgen Dengler*. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Methods

For this study I used all papers (co-)authored by me and available in the Scopus database (https://www.scopus.com) on 1 May 2024. This process allowed me to discuss individual papers without exposing other researchers in an undue manner. Moreover, using the papers of a single author reduces variation resulting from different skills of different authors and from different subject fields in which they work. Of course, the list of co-authors and thus their skills as well as the detailed subject fields still vary, but the latter appear to represent a typical set for vegetation ecologists who publish in the journals of the International Association for Vegetation Science (IAVS).

Data extraction yielded 189 entries, of which four were duplicates, six were from 2024 (i.e. with very limited chance to garner citations, and indeed four were without citations so far) and two were from pre-2003 resulting from unsystematic databasing at that time (one conference abstract, one book review). These 12 entries were excluded, leaving 177 observations to be used in the modelling (Suppl. material 1). These observations stem from 54 journals and two book series, with Journal of Vegetation Science ($n = 25$), Tuexenia (21), Vegetation Classification and Survey (13), Phytocoenologia (12) and Applied Vegetation Science (11) being the most frequent (for more details, see Suppl. material 2).

I used the Field-Weighted Citation Impact (FWCI) as of 1 May 2024, provided by the Scopus database as the measure of scientific impact (dependent variable). FWCI normalises the citations of each paper in the year of the publication and the three following years compared to all papers of a certain year, subject area and article type (e.g. "Article", "Review"). Thus, a FWCI of 1 means that an article was cited as often as the average of all articles in the group; a FWCI of 2 means that it received twice as many citations etc. Unlike the raw citation rates, which are strongly dependent on the time elapsed since the publication, FWCI values are directly comparable between articles published in different years, between reviews and research articles or between different disciplines. Another advantage of the FWCI is that Scopus also provides an analogous measure at the journal level, called Source Normalized Impact per Paper (SNIP), where SNIP is essentially the average of the FWCI values of all articles in the respective period. For one article from 2022 and two articles from 2023 which therefore possessed a FWCI of 0, I inserted half of the minimum of all other FWCI values of that year instead (0.05 and 0.30, respectively) to allow modelling (see below). For readers who are more familiar with Journal Impact Factors (JIFs) from the Web of Science, I calculated the relationship of the two metrics for the year 2022 for those 46 journals that were also included in the Web of Science with linear regression after log-transformation of both variables to meet the assumptions of linear models: $\log_{10}(\text{JIF.2022}) = 0.42 + 1.32 \log_{10}(\text{SNIP.2022})$. This means that a SNIP of 1 corresponds to a JIF of 2.6 and a SNIP of 2 to a JIF of 6.6.

As predictor variables, I used formal and quantitative features of the journal, of the article, its titles and authors, where there is some plausible relationship to citation impact and that could be derived from the data provided by Scopus, or I could easily extract this from pdf's (Table 1). The relevant variables, such as the number of authors, were extracted from the downloaded file from Scopus via text functions in MS Excel, followed by careful manual checking. For the question whether a title contains a dash, all different types of dashes used in Scopus were considered. In cases where Scopus included two language versions in the title field, only that in the language of the article was retained. I considered those articles as "open access" that were categorized as Gold OA, Hybrid Gold OA or Bronze OA in Scopus, while those labelled as Green OA were checked on the journal webpage for free accessibility of the definitive article version. Articles of five journals that are indeed gold or diamond open access, but not or only partly categorized as such in Scopus (Tuexenia, Lazaroa, Preslia, Ecography, Diversity and Distributions) were also assigned to the OA category in the analyses. The FWCI of each article as well as the SNIP.2022 of the journals were taken directly from the Scopus website. One article from a journal whose coverage was discontinued in Scopus and thus did not have a recent SNIP value, received an arbitrary value of half of the minimum of all other SNIPs assigned (0.034). Whether an article was part of a special feature was derived from the pdf's.

Table 1. Variables used in the regression modelling of the 177 articles and some further citation metrics, their value distribution and their handling in the modelling.

Variable	Mean	Min	Max	Modelling
Dependent variable				
FWCI 2024.5	2.94	0.05	32.05	$\log_{10}$
Independent variables (numeric)				
SNIP 2022	1.15	0.03	11.59	$\log_{10}$
Year	2017	2003	2023	
Pages	16.50	1	262	$\log_{10}$
Authors	25.72	1	601	$\log_{10}$
Title characters	94.49	14	209	excluded because of high correlation with Title words
Title words	12.62	1	31	
Independent variables (binary)				
Book chapter	Yes =	3		modelled separately
Open access	Yes =	101		
Special feature	Yes =	63		
English	Yes =	170		
Title with statement*	Yes =	14		
Title with word play**	Yes =	6		
Title with "?"	Yes =	4		
Title with ":"	Yes =	47		
Title with dash	Yes =	27		
Further citation metrics (not used in the modelling)				
Citations	49.02	0	1025	
delta (FWCI vs. SNIP)	1.68	-7.89	30.66	
log-ratio (FWCI vs. SNIP)	0.18	-1.26	1.46	

* "title with statement" means factual statements with a verb, e.g. "Root traits explain plant species distributions along climatic gradients yet challenge the nature of ecological trade-offs"

** "title with word play" means any title where words are used in unusual or metaphoric manner, e.g. "Step(pe) up!" or "Midas touches"

All statistical modelling was done in R version 4.2.2 (R Core Team 2022), assuming a significance threshold of 0.05. The highly skewed variables FWCI, SNIP, Pages and Authors were $\log_{10}$-transformed to achieve approximate normal distribution. The regressions were run with the command 'lm', and the compliance of the final models with the assumptions of linear models was verified via visual inspection of the residual plots (Quinn and Keough 2002). Since books do not have FWCI values in Scopus, the differences in citation rates between book chapters and journal articles were initially tested, while all further modelling was done only for the 174 remaining journal articles. For this purpose, all potential numeric and binary predictor variables of Table 1 were first subjected to a correlation analysis to detect pairs of highly correlated variables. Accordingly, Title. characters was excluded from further modelling as it was highly positively correlated with Title.words ($r = 0.94$). The remaining 12 predictor variables defined the global model, which then was refined stepwise until only significant terms remained (minimal adequate model; see Crawley 2014). With each of the predictor variables of the minimal adequate model, I also conducted simple linear regressions with the log-transformed FWCI as response variable. Since FWCI was modelled on a $\log_{10}$-scale, the raw estimates (given in Table 2) in the following text were back-transformed to linear scale to allow easier interpretation.

Results

The log-transformed FWCI was significantly higher in book chapters than in journal articles ($p = 0.017$; $R^2_{adj.} = 0.026$). The estimate (0.748) suggests that on average my book chapters are cited 5.6 times more often than my journal articles. In the multiple regression for journal articles only, among the 12 predictor variables in the global model, seven remained as significant terms in the minimal adequate model (Table 2).

Table 2. Minimal adequate model to explain the $\log_{10}$-transformed Field-Weighted Citation Impact (FWCI). The estimates for the predictors in the multiple and simple linear regressions as well as the associated $R^2_{adj.}$ values are given. n.s. = non-significant.

	Multiple regression				Simple regressions	
Variable	Estimate	*t* value	*p*-value	$R^2_{adj.}$	Estimate	$R^2_{adj.}$
(Intercept)	0.329	2.009	0.046	0.502		
$\log_{10}$(SNIP 2022)	0.780	7.001	<0.001		0.800	0.268
Special feature	0.158	2.444	0.016			n.s.
Year - 2003	-0.030	-4.620	<0.001			n.s.
$\log_{10}$(Pages)	0.324	2.819	0.005		0.307	0.020
$\log_{10}$(Authors)	0.282	4.067	<0.001		0.416	0.172
Title words	-0.038	-5.692	<0.001		-0.038	0.105
Title with ":"	0.169	2.566	0.011			n.s.

The most influential variable (i.e. the one with the highest absolute *t*-value) in the multiple regression was the log-transformed **SNIP**. The estimate suggests that with each doubling of the SNIP, the FWCI increases on average by 43%. However, in a simple regression SNIP explained only 26.8% of the overall variance in FWCI. Conversely, the minimal adequate model leaving out SNIP explained 31.5% of the variance (not shown).

The **number of title words** had the second-strongest influence in the minimal adequate model. The estimate suggests that each additional word decreases the FWCI by 8.4%, and likewise each word less increases it by 9.1%. Also, the log-transformed **number of authors** was highly significant in the minimal adequate model and was the second-most influential variable among the bivariate models (17.2% explained variance in FWCI). According to the estimate in the minimal adequate model, each doubling of the author numbers would lead to a 13.9% higher FWCI. The **year of publication** had a highly significant negative impact on the FWCI, with an estimated decrease of FWCI per year by 6.7%. By contrast, in the simple regression model year of publication was not significant. The log-transformed **number of pages** was significant, with an estimated increase of the FWCI for each doubling of the page number by 16.1%. The presence of a **colon** (":") in the title had a significant positive impact on the FWCI (+48%) as had the question whether an article was published in a **special feature/special collection** (+44%).

By contrast, the variables **open access** (yes vs. no), **language of the article** (English vs. German) as well as the use of **factual statements**, **questions**, **word plays** or **dashes** in the title had no significant influence on the FWCI in the multiple regression model and thus were not included in the minimal adequate model.

Discussion

Potential mechanisms behind the patterns

Among the tested variables, SNIP was the strongest predictor both in the multiple regression and among the bivariate regressions. It is self-evident that there must be a positive relationship between the FWCI of the articles and the SNIP values of the journals as the latter essentially are the averaged FWCI values of the included articles. That articles in journals with higher SNIP are more cited can be explained by three mechanisms that act together: (1) authors tend to submit their better manuscripts to the better journals; (2) higher-ranked journals likely have the more experienced editors and reviewers who can help more to improve the manuscript than in lower-ranked journals; and (3) publications in higher-ranked journals likely attract more readers as a high SNIP/JIF to many readers suggests high quality. Given all these obvious links, it is somehow astonishing that SNIP explained only a little more than one quarter of the variance in FWCI and thus less than the other formal aspects combined. This is mainly driven by the fact that the citations rates among different articles in the same journal vary dramatically (Figure 1, see Suppl. material 1 and Table 3). For example, the FWCI values of my articles in Vegetation Classifica-

tion and Survey ranged from 0.10 to 18.45 (185-fold difference) and those in Journal of Vegetation Science from 0.30 to 22.82 (76-fold difference, noting that the lowest values in both cases were 0, but these were replaced according to the Methods). A subordinate reason for the relatively low explained variance could be that for simplicity I used the SNIP values of the year 2022 and not of the publication year, thus not accounting for potential systematic changes in the relative positions of journals over the years, which, however, are usually small and gradual. An interesting finding is that the estimate for SNIP was 0.78 in the analysed dataset, whereas across all articles of any authors it should be 1 given that SNIPs essentially are averaged FWCIs. This deviation suggests that for lower-ranked journals my papers are much more cited than the journal average, but this relative difference decreases with increasing journal impact (SNIP). This can be seen also in Table 3 and in Suppl. material 2, if sorted by the difference of FWCI vs. SNIP: The differences are on average highest for relatively low-rank journals and become smaller or even negative for many higher-ranked journals.

Interestingly, the second-most influential predictor was the title length, with articles being on average much more cited when the title is shorter. It is not directly intuitive why title brevity is so influential. Likely, the main reason is that a short title is normally achieved by getting rid of as many non-necessary words as possible. As people find articles mainly via search engines, the title essentially should be a sequence of probable keywords for which people might search ("search engine optimisation"). The top-ranked journal Nature apparently is fully aware of the importance of short titles as their author guidelines strictly forbid any title longer than 75 characters, including spaces (which typically corresponds to 7 to 11 words).

By contrast, the two other numeric indicators, number of authors and number of pages, had a positive effect on citation rates. The particularly strong effect of the number of authors (third-strongest predictor) can be explained by a set of non-exclusive mechanisms. First, a higher number of authors is typically related to larger datasets that allow more comprehensive analyses. Second, if more authors with their experiences are involved in paper preparation, this will likely lead to a higher manuscript quality. Last, a higher number of authors also means that more people (the authors and their networks) are aware of the paper and thus likely to cite it. It is not so obvious why also a greater length of the paper is beneficial. Most likely it is because a greater length allows incorporation of more different subtopics, meaning that the paper contains relevant information for a wider range of other studies.

Among the different binary article typologies, only book chapters vs. journal articles and special features vs.

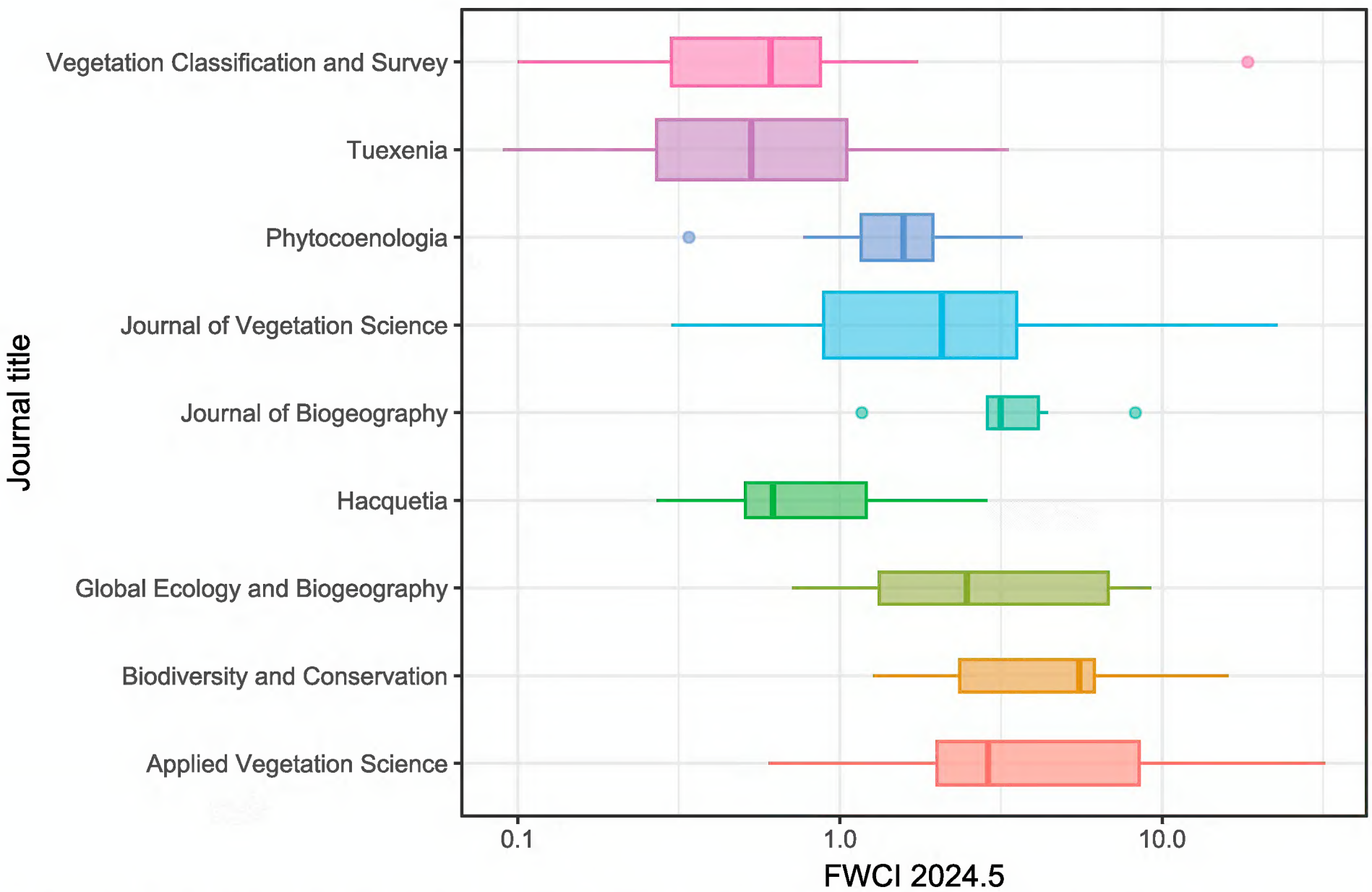

Figure 1. Variation of FWCI values of articles in journals represented by at least five articles in the sample. The height of boxplots is proportional to the number of articles included in the sample. Note that the *x*-axis has a log-scaling. The length of the box-whisker plots indicates that except for Journal of Biogeography, the most-cited article in the sample performs at least 10 times better than the least cited one, while the difference was as big as 185 times in the case of Vegetation Classification and Survey.

regular articles had a positive effect, but not so open access or English language. The unexpectedly much higher citation rate of book chapters compared to journal articles can probably be attributed to the narrow selection of books that are currently covered by Scopus. In my case, these are two "encyclopedias" that provide authoritative mini reviews on the current state of knowledge across a wide range of topics and thus are relevant for many studies as background information. If the coverage of books in Scopus was as wide as for journals, this citation advantage probably would disappear. The citation advantage of articles in special features is not a big surprise. Being part of a special feature automatically increases the visibility as there is usually an editorial that highlights the relevance of each included paper, plus often some additional "advertising" activities. Moreover, editors of special features are specialised in its narrow topic and thus might be able to contribute more to the improvement of the submitted manuscripts than normal editors can in journals where they must handle manuscripts of a much wider range of topics. Surprisingly, publishing OA did not bring any benefit in terms of citation rates. Naively, one would imagine that OA increases the visibility of articles and thus the chance of being cited – and previously there have been some studies that showed such a positive effect (Hajjem et al. 2005), but it is always hard to control for confounding factors. English language probably did not play a significant role in the multiple regression model because most journals publish in one language only and therefore the almost surely existing lower citation rates of German vs. English articles was already accounted for by the journal SNIP. In this dataset, only one journal contained articles in different languages (Tuexenia), and the sample size thus was too small to detect a pattern even if was there.

Among the other characteristics of the article titles beyond the length, only the presence of a colon (":") had a significant positive effect, while using a dash or a word play or phrasing the title as a question or factual statement had no significant effect – despite many authors seem to believe that it is beneficial to do so. In fact, using questions or statements even has an implicit negative effect on citation rates as reformulating a "conventional" title as question or statement requires additional words, while the number of words has a strong negative effect on citation rates. By contrast, the use of colons and dashes allows conveying the same information in a title but with less words, e.g. "Dry grasslands of Southern Europe: Syntaxonomy, management and conservation" instead of "Dry grasslands of Southern Europe with a focus on syntaxonomy, management and conservation". Therefore, it is logical that the use of a colon or dash to separate a subtitle from a title are beneficial for citation rates via the strong effect on title brevity. However, it remains unclear why the colon has an additional strong positive effect while the dash – despite almost identical usage – has not.

Last but not least, there was the surprising result that my citation impact per article highly significantly decreased over the years in the multiple regression model, while the simple regression suggested no change over time. This is unexpected, as one should assume that in this 20-year period, I should have gained experience and now be able to write articles with higher impact than before. Perhaps I did, but it may be that other scientists improved even faster, and this then is reflected in a decrease in mean FWCI per paper – since FWCI values are normalised to the average in the respective research field and year. However, the absence of a change in the bivariate regression points in another direction: I may have improved various things over time, such as targeting higher-impact journals, shorter titles or more co-authors, but these improvements were accounted for already by the other predictors in the model.

Regression model exemplified for this paper

The regression model developed in Table 2 allows one to forecast the FWCI of this paper itself. It has the following parameters: SNIP.2022 = 0.65, SF = 0, Year = 2024, Pages = 9, Authors = 1, Title.words = 4, Title.colon = 0. Inserted into the equation, this would yield a predicted FWCI of 0.504, i.e. below the current average of the journal (SNIP.2022 = 0.65). It should be noted that (a) about half of the variation in the citation rate is not explained by the seven formal variables used in the model and (b) the model is for May 2024 and the then current SNIP and FWCI values. Therefore, if readers should find the content of the paper interesting and useful, it could still become as much cited as average VCS papers or even more.

This estimate helps to explain how different simple choices under my influence as author would have altered the outcome. Originally, I thought of the title "What makes a paper successful?" but abandoned it, when I realised that questions do not improve citation rates but lead to longer titles (in this case: + 1 word). The prediction for this title would be a FWCI of 0.462, i.e. a 8% lower citation rate. If I had chosen to follow the trend to state the main findings in the title, e.g. "Title brevity and article length increase the citation rates of articles", the predicted FWCI would be 0.275, i.e. 45% lower than for the chosen solution. On the other hand, if I had found three more co-authors or expanded the paper with more content to 18 pages, it would likely get more cited (+48% and +25%, respectively).

Limitations

Evidently, the strongest limitation of this study is the small sample size of < 200 articles. Thus, this study cannot (and is not intended to) replace a comprehensive analysis with a much broader dataset. However, since the sample covers a relatively wide range of > 50 journals relevant to vegetation ecologists, the findings still can claim some generality. This is particularly true when focussing on the two strongest predictors (those with the lowest p- and highest R^2-values) after the journal impact (SNIP), i.e. number of authors and number of title words. Actually, the same two variables had

turned out to be highly influential in the same direction in an unpublished study conducted by Meelis Pärtel sometime ago, for all the articles published in Journal of Vegetation Science and Applied Vegetation Science over several years.

Also, the metric of citation impact used here, FWCI, while it was chosen for its obvious advantages over metrics such as the mere citation count, still has limitations. On the Scopus website it is pointed out that the FWCI of an article is less meaningful when its calculation was based on averaging a small group of articles where a single high-impact article could have undue effects. However, this is not the case in the subject areas studied here, each of which is populated by numerous journals, together publishing >> 1000 articles per year. Moreover, the subject area classification by Scopus (ASJC = All Science Journal Classification) as any typology has arbitrary elements. However, these are to some extent levelled out by the fact that most journals are assigned to multiple subject areas; Vegetation Classification and Survey for example to 1110 ("Plant Science"), 1101 ("Agricultural and Biological Sciences (miscellaneous)") and 1105 ("Ecology, Evolution, Behavior and Systematics"). Evidently, assignment to other subject areas would have led to slightly different FWCI values. However, in the current study this potential bias was counteracted by the fact that the journal SNIP is based on exactly the same subject areas as the FWCI of an article.

Recommendations for different stakeholders

Authors

This study underlines that trying to get a certain paper accepted in the journal with the highest possible SNIP or JIF will, if successful, on average lead to higher citation rates, as is in agreement with common sense. However, the study also makes clear that the average impact of the journal only determines slightly more than one quarter of the impact of an article, while the latter should be the focus of authors. This means that it could be more efficient for authors to work on the other formal aspects addressed here, which together have more influence on the article impact than the level of the journal has. For example, instead of trying to publish in a journal with a twice as high journal impact (measured as SNIP), they could shorten their title in a meaningful way by 62%, which probably would cost only a small fraction of the time. Likewise, authors should question the current fashion to formulate the main results in the title as a factual statement, as I could show that by itself it is not beneficial for the impact but leads to a much longer title, resulting in a lower impact (e.g. in the example of the previous section: –45%).

Table 3. The top-5 over- and underperforming papers in the analysed portfolio of 174 journal articles compared to the average citation rates of the respective journals. The ranking was done by absolute differences (delta), while additionally the relative differences are given as ratios and log-ratios. Note that some articles are underperforming relative to the average of the journal in which they were published, but still are overperforming relative to all articles in the subject area and year (i.e. have a FWCI > 1).

Authors	Year	Title	Publication venue	Citations	FWCI 2024.5	SNIP 2022	delta (FWCI vs. SNIP)	ratio (FWCI vs. SNIP)	log-ratio (FWCI vs. SNIP)
Mucina et al.	2016	Vegetation of Europe: hierarchical floristic classification system of vascular plant, bryophyte, lichen, and algal communities	Applied Vegetation Science	1025	32.05	1.389	30.66	23.07	1.363
Tichý et al.	2023	Ellenberg-type indicator values for European vascular plant species	Journal of Vegetation Science	35	22.82	0.901	21.92	25.33	1.404
Dengler et al.	2023	Ecological Indicator Values for Europe (EIVE) 1.0	Vegetation Classification and Survey	22	18.45	0.647	17.80	28.52	1.455
Bruelheide et al.	2018	Global trait–environment relationships of plant communities	Nature Ecology and Evolution	394	20.30	3.989	16.31	5.09	0.707
Wilson et al.	2012	Plant species richness: The world records	Journal of Vegetation Science	609	17.19	0.901	16.29	19.08	1.281
[...]									
Klotz et al.	2022	Plasticity of plant silicon and nitrogen concentrations in response to water regimes varies across temperate grassland species	Functional Ecology	1	0.26	1.645	-1.39	0.16	-0.801
Laughlin et al.	2023	Rooting depth and xylem vulnerability are independent woody plant traits jointly selected by aridity, seasonality, and water table depth	New Phytologist	1	0.62	2.490	-1.87	0.25	-0.604
Vetter et al.	2020	Invader presence disrupts the stabilizing effect of species richness in plant community recovery after drought	Global Change Biology	18	1.04	3.007	-1.97	0.35	-0.461
Jandt et al.	2022a	ReSurveyGermany: Vegetation-plot time-series over the past hundred years in Germany	Scientific Data	5	0.58	2.887	-2.31	0.20	-0.697
Jandt et al.	2022b	More losses than gains during one century of plant biodiversity change in Germany	Nature	27	3.70	11.591	-7.89	0.32	-0.496

Editors, reviewers and publishers

Most editors probably would agree that their job is to select those articles that are not only topic-wise but also impact-wise matching the journal, i.e. avoid articles that will become much less cited than the journal average. This study suggests that editors are not very good in this selection as the variation of article impact within the individual journals is extreme (see Suppl. material 2). Taking only the small subset of articles co-authored by me, the variation of FWCI in those five journals where I had at least 10 papers ranged from 11- to 185-fold (and similarly for all journals with at least five papers, see Figure 1), meaning that the range of all articles in these journals must be even wider. While editors can be happy if authors submit articles that then perform much better than the average of the other articles in the journal, one could argue that they make a misjudgement if they accept articles that are far less cited than expected for an average article in that journal. The most extreme examples of negative and positive mismatches among my portfolio are shown in Table 3.

I hope that this Forum contribution can raise the awareness among editors that currently they are often not doing a particularly good service to their journals in deciding which manuscripts to accept or reject, at least not from the perspective of scientific impact. I believe that editors could and should be trained much better to forecast the potential scientific impact of submitted manuscripts – which evidently concerns not only the 31.5% of variance explained just by formal issues discussed here, but also the 49.8% of (probably mostly content-wise issues) not addressed here. This refers both to avoid rejections of potential high-impact papers, and acceptances of papers that likely will be much less attractive than average articles in that journal. For example, the article by Bruelheide et al. (2018) from Table 3 was originally submitted to Nature but rejected – and now has citation rates almost twice as high as average Nature papers (FWCI = 20.30 vs. SNIP.2022(Nature) = 11.59). From Nature perspective, this editor decision was thus not beneficial.

Another simple issue that journals could ask themselves is whether the strict upper thresholds for article length defined in many author guidelines are still appropriate, given that longer papers receive significantly more citations after taking all other aspects into consideration. Page limits made sense in the old times when articles were still printed on paper and journal issues sent by mail, i.e. each additional page came with substantial additional costs, but in times of electronic publishing when a few pages more cost hardly anything, this does not appear wise. But of course, editors should only accept longer articles when the additional pages are justified by the content.

Science funders and universities

This study calls into question several widespread practices of science funders and universities.

In many countries, researchers are strongly pushed to publish their results in "high-rank" journals, often defined as the first and second quarter of JIFs in the Web of Science database. I consider this practice clearly unethical. First, it removes the decision on what is valuable science from scientists and puts it into the hands of a commercial enterprise (Clarivate) and their arbitrary and intransparent decisions as to which journals to include in their database at all.

Second and perhaps more importantly, the variation of citation rates within most (if not all) of the journals is so extreme that it is arbitrary and unfair to assess the impact of an article by the average impact of all articles in that journal. Why should the Nature article by Jandt et al. (2022) (Q1 in Web of Science) be more valuable than the Ecological Indicator Values for Europe (Dengler et al. 2023) (not included in the Journal Citation Report of Web of Science so far and, if included, probably Q3) despite the latter having a five times higher citation impact (both are far above the average citation rates in their disciplines; Suppl. material 1)? This misuse of journal-based metrics for the assessment of scientists has been repeatedly criticised (CoARA 2022), still it seems to be the prevailing practice in many countries as I hear from my colleagues abroad. In the past, there might have been some sense in using a ranking based on JIF, CiteScore or SNIP of the journals, when there were no meaningful article-based metrics. However, now we have the FWCI, which directly measures the normalised impact of each article, and is readily available not only for the journals in the Web of Science but for the much wider selection of journals in the Scopus database, and becomes available latest one year after publication (see also Dengler et al. 2024).

Thirdly, this study calls into question one of the major motivations for the OA movement: to make scientific results better accessible (BOAI 2002; Tennant et al. 2016). In the earlier days of OA, studies claimed to have found significant citation advantages (Hajjem et al. 2005). However, in the dataset analysed here, being published behind a paywall apparently did not create any relevant impediment for other scientists to access the articles. Actually, universities have still subscribed to thousands of non-OA journals, which makes the access easy. But even if not, it is not a real challenge to get an article – just write an e-mail to the authors and nearly always they will be happy to send you a copy. This has become even easier today when you can use Research Gate (https://www.researchgate.net) to find an article and, when it is not open access, just press a button and an automated email to the authors is generated. I can say that in my life as a scientist I have received all articles I requested, but I (or my university) have never paid any article fee for that – despite being at small- to medium-size universities that had subscription access only to a limited number of journals. Therefore, one might ask whether it is a wise allocation of public money to spend millions of Euros or dollars on OA if this does not lead to substantially better access, in particular if most of the money just ends up as profit in the pockets of a few mega-publishing houses, whether predatory or not (Dengler 2023).

Among the national science funders who did and still do push OA publishing massively is the Swiss National Science Foundation (SNSF), which recently started to admit that there are some negative side effects. In consequence of that, they stopped paying OA fees for articles in special features (SNSF 2023) instead of stopping payments to predatory publishers and journals. While it is true that many predatory journals have a particularly high fraction of articles in special features (Hanson et al. 2023), it appears unethical to ban the financial support of all special features, even those in serious journals where they typically contain the most prestigious content (Ibrahim et al. 2024). The latter assumption has been ascertained by this study that found 44% higher citation impact for special feature contributions, all other things being equal (and there is no predatory journal in my publication portfolio).

Bibliometric databases

It should be highlighted that this whole study became only possible by the Scopus database providing the matching pair of normalised citation indices, both for the journals (SNIP) and the individual articles (FWCI). The normalisation makes studies across subject areas with different citation practices and across years (with different numbers of articles, e.g. the publication peak in the COVID-19 years: Dengler 2023) possible and is similarly available in the Web of Science under the name Journal Citation Indicator (JCI). In this respect, both SNIP and JCI are much more informative indicators than the more widely used CiteScore (Scopus) and JIF (Web of Science) (see also Dengler et al. 2024). However, only Scopus currently provides a matching meaningful indicator at the article level, which is the one thing that is really needed in fair impact assessment of scientists. Unfortunately, Scopus still does not advertise the FWCI as prominently as they could, and it is still not particularly user-friendly to get the data. Unlike the absolute numbers of citations, which can be automatically downloaded for a larger list of articles, the FWCI currently must be retrieved manually for each article separately. Smaller issues with the use of the SNIP and FWCI concern the fact that book chapters currently do not have a SNIP and that the article categorizations used for the normalisation are not always consistently applied (e.g. some editorials currently are coded as editorial, others as reviews), which has an influence on the SNIP and FWCI calculation in a smaller fraction of cases. However, already now, according to my experience, the combination of SNIP and FWCI offers more informative analyses than any of the indices available in the Web of Science do.

Conclusions

I would like to emphasize that authors, reviewers, editors and science funders should primarily aim for high-quality science. However, I have shown here that the impact of one specific paper is not only defined by its scientific qualities, but to a non-negligible part also by simple formal aspects. As author, it is worth being aware of these mechanisms and take advantage of them to make your own high-quality papers as impactful as they can be. Likewise, reviewers and editors could use this empirical knowledge to give better advice to their authors. I thus hope that this contribution opens a wider discussion on the relevance of formal aspects for the scientific impact of articles. Evidently, this was just an example study based on a small sample from a single vegetation ecologist. However, the results largely coincide with an unpublished study by Meelis Pärtel who several years ago analysed the publication output of Journal of Vegetation Science and Applied Vegetation Science over several years. Hopefully this Forum Paper will spur much more comprehensive follow-up studies across multiple authors and disciplines to test how general the reported patterns are.

Data availability

All data used are provided in the Supplementary materials.

Acknowledgements

I would like to thank Meelis Pärtel who, several years ago when he was the Chair of the Chief Editors of *Applied Vegetation Science* and *Journal of Vegetation Science*, conducted a similar study on articles published in these two journals. His study came to similar conclusions as this one, but unfortunately was never published. This motivated me to finally get something citable on the topic. Many thanks to François Gillet and Idoia Biurrun who made very useful suggestions to a former versions of the manuscript. Further, I am grateful to Stephen Bell for linguistic revision of this article.

References

BOAI (2002) Budapest Open Access Initiative. https://www.budapestopenaccessinitiative.org/read/ [accessed 30 Jul 2023]

Bruelheide H, Dengler J, Purschke O, Lenoir J, Jiménez-Alfaro B, Hennekens SM, Botta-Dukát Z, Chytrý M, Field R, … Jandt U (2018) Global trait–environment relationships of plant communities. Nature Ecology and Evolution 2: 1906–1917. https://doi.org/10.1038/s41559-018-0699-8

Cargill M, O'Connor P (2009) Writing scientific research articles – strategy and steps. Wiley-Blackwell, Chichester, UK.

CoARA (2022) Agreement on reforming research assessment. 20 July 2022. Coalition for Advancing Research Assessment. https://coara.eu/ [accessed 5 May 2024]

Crawley MJ (2014) Statistics – An introduction using R. 2nd ed. John Wiley & Sons, Chichester, UK.

Dengler J (2023) Priorities in journal selection for authors, reviewers, editors, librarians and science funders. Vegetation Classification and Survey 4: 219–229. https://doi.org/10.3897/VCS.110296

Dengler J, Jansen F, Chusova O, Hüllbusch E, Nobis MP, Van Meerbeek K, Axmanová I, Bruun HH, Chytrý M, … Gillet F (2023) Ecological Indicator Values for Europe (EIVE) 1.0. Vegetation Classification and Survey 4: 7–29. https://doi.org/10.3897/VCS.98324

Dengler J, Biurrun I, Jansen F, Willner W (2024) Vegetation Classification and Survey is performing well. Vegetation Classification and Survey 5: 1–10. https://doi.org/10.3897/VCS.118454

Gustavii B (2008) How to write and illustrate a scientific paper. 2nd ed. Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/CBO9780511808272

Hajjem C, Harnad S, Gingras Y (2005) Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. IEEE Data Engineering Bulletin 28: 39–47.

Hanson MA, Barreiro PG, Crossetto P, Brockington D (2023) The strain of scientific publishing. arXiv 2309.15884v1. https://doi.org/10.48550/arXiv.2309.15884

Ibrahim S, Zhang Y, Reusser K (2024) AI-generated rat genitalia: Swiss publisher of scientific journal under pressure. SWI swissinfo.ch [accessed 6 May 2024]

Jandt U, Bruelheide U, Berg C, Bernhardt-Römermann M, Blüml V, Bode F, Dengler J, Diekmann H, Dierschke H, … Wulf M (2022a) ReSurveyGermany: Vegetation-plot time-series over the past hundred years in Germany. Scientific Data 9: e631. https://doi.org/10.1038/s41597-022-01688-6

Jandt U, Bruelheide H, Jansen F, Bonn A, Grescho V, Klenke R, Sabatini FM, Bernhardt-Römermann M, Blüml V, … Wulf M (2022b) More losses than gains during one century of plant biodiversity change in Germany. Nature 611: 512–518. https://doi.org/10.1038/s41586-022-05320-w

Klotz M, Schaller J, Feldhaar H, Dengler J, Gebauer G, Aas G, Weissflog A, Engelbrecht B (2022) Plasticity of plant silicon and nitrogen concentrations in response to water regimes varies across temperate grassland species. Functional Ecology 36: 3211–3222. https://doi.org/10.1111/1365-2435.14225

Laughlin D, Siefert A, Fleri J, Tumber-Dávila S, Hammond W, Sabatini F, Damasceno G, Aubin I, Field R, … Bruelheide H (2023) Rooting depth and xylem vulnerability are independent woody plant traits jointly selected by aridity, seasonality, and water table depth. New Phytologist 240: 1774–1787. https://doi.org/10.1111/nph.19276

Mucina L, Bültmann H, Dierßen K, Theurillat JP, Raus T, Čarni A, Šumberová K, Willner W, Dengler J, … Tichý L (2016) Vegetation of Europe: Hierarchical floristic classification system of vascular plant, bryophyte, lichen, and algal communities. Applied Vegetation Science 19(suppl. 1): 3–264. https://doi.org/10.1111/avsc.12257

Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/CBO9780511806384

R Core Team (2022) R: A language and environment for statistical computing. Version 4.2.2. R Foundation for Statistical Computing, Vienna, AT. https://www.R-project.org/

SNSF (2023) The SNSF is no longer funding Open Access articles in special issues. https://www.snf.ch/en/g2ICvujLDm9ZAU8d/news/the-snsf-is-no-longer-funding-open-access-articles-in-special-issues [accessed 6 May 2024]

Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink CHJ (2016) The academic, economic and societal impacts of Open Access: an evidence-based review. F1000Research 5: e632. https://doi.org/10.12688/f1000research.8460.3

Tichý L, Axmanová I, Dengler J, Guarino R, Jansen F, Midolo G, Nobis MP, Van Meerbeek K, Aćić S, … Chytrý M (2023) Ellenberg-type indicator values for European vascular plant species. Journal of Vegetation Science 34: e13168. https://doi.org/10.1111/jvs.13168

Vetter V, Kreyling J, Dengler J, Apostolova I, Arfin Khan MAS, Berauer BJ, Berwaers S, De Boeck H, Nijs I, … Jentsch A (2020) Invader presence disrupts the stabilizing effect of species richness in plant community recovery after drought. Global Change Biology 26: 3539–3551. https://doi.org/10.1111/gcb.15025

Wilson JB, Peet RK, Dengler J, Pärtel M (2012) Plant species richness: the world records. Journal of Vegetation Science 23: 796–802. https://doi.org/10.1111/j.1654-1103.2012.01400.x

E-mail and ORCID

Jürgen Dengler (Corresponding author, juergen.dengler@zhaw.ch), ORCID: https://orcid.org/0000-0003-3221-660X

Supplementary material

Supplementary material 1
Overview of the 177 articles analysed, broken down to publication venue with journal- and article-based citation metrics and the analysed predictor variables (*.xlsx).
Link: https://doi.org/10.3897/VCS.126956.suppl1

Supplementary material 2
Overview of the 54 journals and two book series included in the analysis, with journal-based and article-based metrics and their relationships (*.xlsx).
Link: https://doi.org/10.3897/VCS.126956.suppl2